

Graph decomposition and parity

Bobby DeMarco ^{*} and Amanda Redlich [†]

November 9, 2012

Abstract

Motivated by a recent extension of the zero-one law by Kolaitis and Kopparty, we study the distribution of disconnected subgraphs in the random graph $G(n, p)$. We use graph decompositions to give a sufficient condition for the distribution to tend to uniform modulo q . We determine the asymptotic distribution of all two-component graphs in $G(n, p)$ for all q , and we give infinite families of many-component graphs with a uniform asymptotic distribution for all q . We also prove a negative result, that no simple proof of uniform asymptotic distribution for arbitrary graphs exists.

1 Introduction

A recent paper by Kolaitis and Kopparty [2] gives an extension of the zero-one law which holds for first-order logic with a parity operator. The keystone of their proof is that the number of copies of any connected subgraph is asymptotically uniformly distributed modulo q for any q in the random graph $G(n, p)$. In this paper we will only be speaking of asymptotic distributions, so when discussing distributions we will remove the word asymptotic for brevity.

Here we study the distribution of the number of copies of *disconnected* subgraphs modulo q . We give a sufficient condition for a disconnected subgraph to be uniformly distributed. We completely characterize the distributions of two-component subgraphs. We give infinitely large families of subgraphs of three or more components that are uniformly distributed.

In analyzing subgraph distributions, we developed the concepts of *unique composition* and *decomposition*. These concepts are related to determining when several subgraphs may be combined to create a larger graph with certain uniqueness properties. There are obvious connections to the reconstruction conjecture (see [1] for a summary), which asks when the subgraphs of a larger graph have slightly different uniqueness properties. In this paper we

^{*} Supported by the U.S. Department of Homeland Security under Grant Award Number 2007-ST-104-000006.

[†] Supported by the National Science Foundation under Award No. 1004382.

give an algorithm for uniquely composing any two subgraphs, and certain families of three or more subgraphs. We also show no generic recursive composition algorithm exists.

The paper is structured as follows. In the second section we analyze the relationship between connected and disconnected subgraph counts. An inclusion-exclusion argument, together with some ideas about set partitions and graph decompositions, leads to a formula for the count of disconnected subgraphs. In the third section we use this formula to give specific conditions for a disconnected graph to be uniformly distributed.

We show these conditions are satisfied for almost all two-component graphs in the fourth section. We give an explicit construction for all satisfying graphs. We also calculate the distribution for all two-component graphs that do not satisfy these conditions. We then give some examples of infinite families of three or more component graphs that are satisfying. We conclude this section with a negative result, showing no simple algorithm exists to show a generic graph is satisfying. In the last section, we discuss areas of further research.

2 Counting copies

In this section we give an exact formula for the number of unlabeled copies of any disconnected graph $A = \sqcup_{i=1}^k G_i$ in a host graph F . This formula is given in terms of the number of copies of various connected graphs H , and their relationship to the original graph A . Although the formula appears complex, the reasoning behind it is simple. The main idea is that each copy of A is the product of copies of G_1, G_2, \dots, G_k . Interactions between copies of G_i and G_j lead to an overcount; the formula uses the principle of inclusion and exclusion to correct for this.

The simplest case is when $A = G_1 \sqcup G_2$. For example, let $G_1 = C_3$ and $G_2 = C_4$. Given fixed host graph F , let $N(A)$ be the number of unlabeled copies of A in F (we may also mean the number of unlabeled copies of A in an instance of the random graph $G(n, p)$; this will be clear from context). Consider H_i as illustrated below. For ease of discussion, here and throughout the paper, we label the illustrated vertices. However, the graphs themselves are unlabeled.

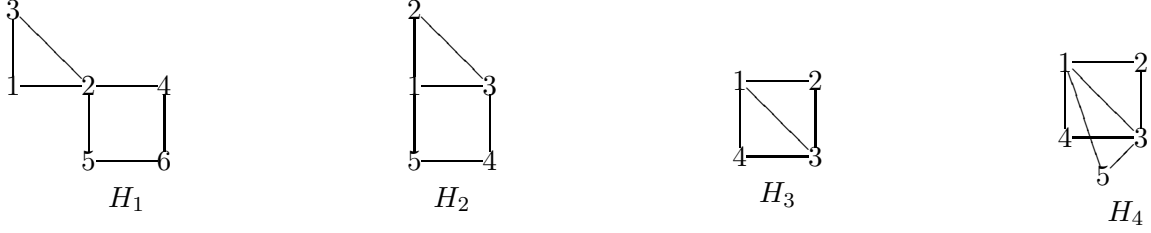
We have

$$N(A) = N(C_3)N(C_4) - N(H_1) - N(H_2) - 2N(H_3) - 3N(H_4).$$

That is because the total number of disconnected C_3, C_4 pairs is the total number of C_3, C_4 pairs minus the number of connected pairs. While H_1 and H_2 each correspond to exactly one connected pair, H_3 is counted twice; once when $C_3 = \{1, 2, 3\}$ and once when $C_3 = \{1, 3, 4\}$. Similarly, H_4 is counted three times: $C_3 = \{1, 2, 3\}$ or $\{1, 3, 4\}$ or $\{1, 3, 5\}$.

We now formalize this “gluing” idea.

Definition 1. A tuple $(G_1, G_2, \dots, G_k, H, H_1, H_2, \dots, H_k)$ is a *gluing* of $G_1 \dots G_k$ if



- H is a connected graph
- H_1, \dots, H_k are subgraphs of H
- $H_i \sim G_i$ for all $i \in [k]$
- $\cup_{i=1}^k E(H_i) = E(H)$.

We occasionally refer to H itself as a gluing. The tuple (H_1, \dots, H_k) is a *decomposition* of H . If there exists only one tuple H_1, \dots, H_k such that $(G_1, \dots, G_k, H, H_1, \dots, H_k)$ is a gluing, we say that (G_1, \dots, G_k, H) is *uniquely decomposable*. In this case, we say that $H[S]$ is the unique subgraph of H induced by $\{H_i\}_{i \in S}$. We occasionally say that H itself is uniquely decomposable, if G_1, \dots, G_k are clear from context. If there exists an H such that (G_1, \dots, G_k, H) is uniquely decomposable, we say that $\{G_1, \dots, G_k\}$ is *uniquely composable*.

We often want to count gluings and decompositions.

Definition 2. Given G_1, \dots, G_k , $s(H)$ is the number of tuples (H_1, \dots, H_k) such that $(G_1, \dots, G_k, H, H_1, \dots, H_k)$ is a gluing. The set of gluings \mathbf{H} is the family of graphs H such that $s(H) \neq 0$.

Then using the above notation, we have:

Theorem 3.

$$N(A) = N(G_1)N(G_2) - \sum_{H \in \mathbf{H}} s(H)N(H)$$

Proof. The number of copies of A is the number of G_1, G_2 pairs overall, less the number of G_1, G_2 pairs that intersect. Each intersecting G_1, G_2 pair corresponds to a decomposition of a copy of some H , so the total number of intersecting pairs is $\sum_{H \in \mathbf{H}} s(H)N(H)$. \square

It is tempting to generalize this to the three-or-more component case as

$$N(A) \text{“} = \text{”} \prod_{i=1}^k N(G_i) - \sum_{H \in \mathbf{H}} s(H)N(H),$$

but the truth is more complicated. Along with H that may be decomposed into G_1, \dots, G_k , we must consider H that are decomposable into any subset of G_i . For example, if A has

components G_1, G_2, G_3 then we must be concerned with gluings of the forms (G_1, G_2, H, H_1, H_2) , (G_1, G_3, H, H_1, H_3) , (G_2, G_3, H, H_2, H_3) , and $(G_1, G_2, G_3, H, H_1, H_2, H_3)$. In order to deal with this complication, we define some new terms. First, some notation about partitions.

Definition 4. Consider the partitions of $[k]$ under partial ordering by refinement, where we use π and ρ for partitions of $[k]$ and say $\pi < \rho$ if π is a refinement of ρ . Let $\pi(T)$ be the family of blocks in π such that $\cup_{S \in \pi(T)} S = T$. For example, if $k = 4$, $\pi = \{\{12\}\{3\}\{4\}\}$, and $\rho = \{\{123\}\{4\}\}$, then $\pi(\{123\}) = \{\{12\}\{3\}\}$.

Now we use partitions to classify gluings. Each connected component corresponds to a set in a partition, and the family of gluings is broken into sub-families according to the partitions they generate.

Definition 5. Let \mathbf{H}_π be the family of graphs $H = \sqcup_{S \in \pi} H_S$ where each H_S may be decomposed (not necessarily uniquely) into $\{H_i\}_{i \in S}$. For example, $\mathbf{H}_0 = A$ and any $H \in \mathbf{H}_1$ is a connected graph.

We also count possible decompositions of gluings. Since we are now considering a broader range of gluings, we must add a subscript to clarify which graphs are being glued.

Definition 6. Given G_1, \dots, G_k and $S \subseteq [k]$, let $s_S(G')$ be the number of ways the graph G' may be decomposed into copies of $\{G_i\}_{i \in S}$.

Finally, it may be useful to discuss decompositions into graphs other than our original components G_i .

Definition 7. Given a graph H_S for each $S \in \pi$, $s_{\pi(T)}(G')$ is the number of ways G' may be split into $\{H_S\}_{S \in \pi(T)}$. If there exists some $\{H_S\}_{S \in \pi}$ such that $s_{\pi([k])}(G') \neq 0$, say that G' is *compatible* with π .

This notation allows us to give a more general recursion for the number of copies of a graph with an arbitrary number of connected components. Although the notation is daunting, it is a simple generalization of the ideas in the two-component case.

Theorem 8. For a graph $A = \sqcup_{i=1}^k G_i$,

$$\begin{aligned} N(A) &= \prod_{i=1}^k N(G_i) - \sum_{\mathbf{0} < \pi \leq [k]} \sum_{H \in \mathbf{H}_\pi} N(\sqcup_{S \in \pi} H_S) \prod_{S \in \pi} s_S(H_S) \\ &= \prod_{i=1}^k N(G_i) - \sum_{\mathbf{0} < \pi \leq [k]} \sum_{H \in \mathbf{H}_\pi} \prod_{S \in \pi} s_S(H_S) \left(\prod_{S \in \pi} N(H_S) - \sum_{\rho > \pi} \sum_{J \in \mathbf{H}_\rho} \prod_{T \in \rho} s_{\pi(T)}(J_T) N(\sqcup_{T \in \rho} J_T) \right) \end{aligned}$$

Proof. Now that we have the proper definitions, the proof is short. As usual, we count the number of copies of A by finding the product of copies of its components, then subtract the overcount. The “overcounted” graphs are those in which at least two G_i intersect with each other, i.e. those corresponding to a non- $\mathbf{0}$ partition. Therefore we have the first line of the equation.

To see why the second line is true, simply apply the first equation to each $N(\sqcup_{S \in \pi} H_S)$ term individually. Now the relevant partitions are those of which π is a refinement, and the decompositions are not into G_i but instead H_S . \square

This theorem gives a recursive algorithm for calculating $N(A)$ for an arbitrary host graph. With sufficient computing power, then, we could use it to calculate $N(A)$ for the random graph directly. This theorem can be used as a starting point that will allow us to give explicit counts of a family of graphs, as well as a sufficient condition for graphs to have certain distributions. The first step is to expand the recursion to get a simpler formula.

Lemma 9. *For any graph $A = \sqcup_{i=1}^k G_i$, there exist integers $f_A(H)$ for every $H \in \cup_{\pi \leq [k]} \mathbf{H}_\pi$ such that*

$$N(A) = \prod_{i=1}^k N(G_i) - \sum_{\pi \leq [k]} \sum_{H \in \mathbf{H}_\pi} \prod_{S \in \pi} N(H_S) f_A(H)$$

Note that f_A is uniquely determined; this follows from there being no way to write the number of copies of any connected graph in terms of the number of copies of other connected graphs.

3 Distribution of copies

As mentioned in the introduction, [2] proves that, for any constants p and q , any $i < q$, and any connected graph G_0 , the probability of $G(n, p)$ having i copies of G_0 modulo q tends to $1/q$ as n tends to infinity. That is, the distribution of connected subgraphs in the random graph tends to uniform modulo q . We give exact distributions for the number of copies of any disconnected G_0 in $G(n, p)$ modulo q in this section by combining the formulas of the previous section with these results on connected graphs.

The previous section gives exact expressions for the number of copies of a disconnected graph in a particular subgraph. The formulas are often difficult to implement. However, since our goal is the distribution of subgraphs, rather than exact counts, the preceding formulas are enough. To study the distributions of disconnected graphs, we first recall Theorem 3.2 in [2], which we restate here:

Theorem 10. *For any $q > 1$ and $p \in (0, 1)$, and any family of distinct finite connected graphs F_1, \dots, F_l , the distribution of $(N(F_1), \dots, N(F_l))$ modulo q is $2^{-\Omega(n)}$ close to uniform over $[q]^l$.*

Combining Theorem 10 and Lemma 9 produces the following corollary

Corollary 11. *Given $A = \sqcup_{i=1}^k G_i$, if there exists some $H \in \mathbf{H}_1$ such that $f_A(H)$ is relatively prime to q and H is not a gluing of $\{G_i\}_{i \in S}$ for any $S \subset [k]$, then the distribution of $N(A)$ in $G(n, p)$ is $2^{-\Omega(n)}$ -close to uniform modulo q .*

The rest of this section is concerned with finding, for a given A , an H such that $f_A(H) = \pm 1$. This is enough to show that, for any q , $N(A)$ is distributed uniformly modulo q .

Definition 12. Given $A = \sqcup_{i=1}^k G_i$ and a uniquely decomposable gluing H with decomposition (H_1, \dots, H_k) , the *structure graph* of H , denoted by $T(H)$, is the graph whose vertices are $[k]$ and edges are pairs i, j such that $H_i \cap H_j$ is non-empty.

Definition 13. Given $A = \sqcup_{i=1}^k G_i$, a gluing H is *tree-like* if it is uniquely decomposable and its structure graph, $T(H)$, is a tree.

We now show

Theorem 14. *For any graph $A = \sqcup_{i=1}^k G_i$ and tree-like gluing H , $f_A(H) = (-1)^{k-1}$.*

Proof. The proof is by strong induction. When $k = 2$ the statement follows from Theorem 3: when $k = 2$ any uniquely decomposable gluing is tree-like.

Now consider $k \geq 3$. By the first line of Theorem 8,

$$N(A) = \prod_{i=1}^k N(G_i) - \sum_{0 < \pi \leq [k]} \sum_{J \in \mathbf{H}_\pi} N(\sqcup_{S \in \pi} J_S) \prod_{S \in \pi} s_S(J_S).$$

The induction hypothesis applied to $\sqcup_{S \in \pi} J_S$ implies that

$$f_A(H) = - \sum_{0 < \pi \leq [k]} \sum_{J \in \mathbf{H}_\pi} f_J(H) \prod_{S \in \pi} s_S(J_S) \quad (1)$$

Note that for any J with $f_J(H) \neq 0$, the unique decomposability of H gives that there is exactly one $0 < \pi \leq [k]$ such that $J \in \mathbf{H}_\pi$. Let us call this partition $\pi(J, H)$. We will show

Claim 15. For any r -component J , if $f_J(H) \neq 0$, then $J_S = H[S]$ for all $S \in \pi(J, H)$ and $f_J(H) = -1^{r-1}$.

Claim 16. If $f_J(H) \neq 0$ then $\prod_{S \in \pi(J, H)} s_S(J_S) = 1$.

Claim 17. For $0 < \pi \leq [k]$ the number of $J \in \mathbf{H}_\pi$ such that $J_S = H[S] \forall S \in \pi$ is one if H is compatible with π and zero otherwise. If H is compatible with π , call π H -good.

Claim 18. There are $\binom{k-1}{r-1}$ H -good partitions π consisting of r sets.

Combining Claims 15-18 with (1) we immediately have (think of r as the number of components of J , or equivalently number of sets in the partition π related to J)

$$\begin{aligned} f_A(H) &= - \sum_{r=1}^{k-1} \binom{k-1}{r-1} (-1)^{r-1} \\ &= -1^{k-1}. \end{aligned}$$

Proof of Claim 15 The first part of the claim holds because H is uniquely decomposable. Note that H is a tree-like gluing of the components of J . Thus the second part of the claim is an application of our inductive hypothesis.

Proof of Claim 16 This follows from the unique decomposability of H .

Proof of Claim 17 J_S must be a connected graph. So if $H[S]$ is connected for all $S \in \pi$ then $J = \cup_{S \in \pi} H[S]$ is clearly the only $J \in \mathbf{H}_\pi$ such that $J_S = H[S] \forall S \in \pi$. If $H[S]$ is disconnected, then H was not compatible with π . Thus there are zero such graphs.

Proof of Claim 18 Consider the natural mapping from a partition $0 < \pi \leq [k]$ to the set $E(T(H) \setminus F)$ where $F = \cup_{S \in \pi} H[S]$. This mapping defines a bijection from the H -good partitions π consisting of r distinct sets and the set of subgraphs of $T(H)$ with $r-1$ edges. \square

4 Specific examples

Here we give some applications of the theorems of the previous section. We begin with a complete characterization of the distribution of all two-component graphs, together with explicit constructions. We then give several families of graphs that have path-like gluings, and therefore by Theorem ?? are uniformly distributed. Finally, we show that there does not exist a generic construction for all disconnected graphs.

4.1 Two component graphs

Any uniquely decomposable gluing of two graphs must be tree-like. So one way to show that $N(A)$ is uniformly distributed for some two component graph A would be to give a construction of a uniquely-decomposable H . In fact, such a construction exists.

Theorem 19. *If $G_1 \neq G_2$, neither G_1 nor G_2 is a single vertex, $\{G_1, G_2\} \neq \{P_1, P_2\}$ and $\{G_1, G_2\} \neq \{P_1, P_3\}$, there exists a graph H such that (G_1, G_2, H, H_1, H_2) is a uniquely decomposable gluing and $H \neq G_1, G_2$. Furthermore, H may be constructed explicitly.*

In order to describe the construction of H , we define a few new terms. H will be created by taking two graphs and “gluing” them together.

Definition 20. Given G_1 and G_2 and vertices $v_1 \in V(G_1)$ and $v_2 \in V(G_2)$, to glue v_1 and v_2 to create a new graph H_v is the natural identification:

$$V(H_v) = (V(G_1) \setminus \{v_1\}) \cup (V(G_2) \setminus \{v_2\}) \cup \{v\}$$

and

$$E(H_v) = \{\{x, y\} \mid \{x, y\} \in E(G_1) \cup E(G_2)\} \cup \{\{x, v\} \mid \{x, v_1\} \in E(G_1)\} \cup \{\{v, y\} \mid \{v_2, y\} \in E(G_2)\}.$$

Given G_1 and G_2 and edges $\{u_1, v_1\} = e_1 \in G_1$ and $\{u_2, v_2\} = e_2 \in G_2$, to glue e_1 and e_2 to create a new graph H_e is the natural identification:

$$V(H_e) = (V(G_1) \setminus \{u_1, v_1\}) \cup (V(G_2) \setminus \{u_2, v_2\}) \cup \{v\}$$

and

$$E(H_e) = \{\{x, y\} \mid \{x, y\} \in E(G_1) \cup E(G_2)\} \cup \{\{x, u\} \mid \{x, u_1\} \in E(G_1)\} \cup \{\{x, u\} \mid \{x, u_2\} \in E(G_2)\} \\ \cup \{\{x, v\} \mid \{x, v_1\} \in E(G_1)\} \cup \{\{x, v_2, y\} \mid \{x, v_2\} \in E(G_2)\}.$$

Another difficulty is deciding where to glue two graphs. To describe gluing locations, we consider the underlying connectivity structure of each graph. We say the *block degree* $b_G(v)$ of a vertex $v \in G$ is the number of components generated by the removal of v , i.e. $b(v) = (\text{number of components of } G - v) - (\text{number of components of } G)$. Note that $b_G(v) > 0$ if and only if v is a cut vertex. So every connected graph has at least two vertices of block degree 0, which we will call *block-leaves*. Let $B(G) = \max_{v \in V(G)} b_G(v)$.

Throughout the following discussion we let H refer to the graph created by gluing together G_1 and G_2 at either v_1 and v_2 , or e_1 and e_2 , as discussed in Definition 19. H_1, H_2 will be an arbitrary decomposition of H . That is, H_i may be the original graph G_i , or it may be a different image of G_i in H . Note that, if H is formed by gluing at a vertex, then $H_1 \cap H_2$ is a single vertex. Similarly, if H is formed by gluing at an edge, $H_1 \cap H_2$ is a single edge. We begin with a few observations about the block degree.

Observation 21. If H is made by gluing together G_1 and G_2 at a vertex, then in any decomposition H_1, H_2 , with $z = H_1 \cap H_2$, for all $x \neq z \in H_1$,

$$b_{H_1}(x) = b_H(x)$$

and for all $x \in H_2$,

$$b_{H_2}(x) = b_H(x).$$

Furthermore, $b_{H_1}(z), b_{H_2}(z) \leq b_H(z)$.

Observation 22. If H is formed by gluing together G_1 and G_2 at an edge, then in any decomposition H_1, H_2 , for all $x \notin H_1 \cap H_2$,

$$b_{H_1}(x) = b_H(x)$$

and for all $x \in H_2$,

$$b_{H_2}(x) = b_H(x).$$

For $x \in H_1 \cap H_2$, $b_{H_1}(x), b_{H_2}(x) \leq b_H(v)$.

With these definitions and observations in hand, we begin the proof of Theorem 19.

Proof. Without loss of generality, we may assume $B(G_2) \geq B(G_1)$ and, if $B(G_2) = B(G_1)$, then $|V(G_1)| \leq |V(G_2)|$. Let S_i be the set of vertices in G_i of block-degree $B(G_i)$. We split graph pairs into five cases, according to their block degrees, and give a construction for each case.

Case A: $B(G_2) > B(G_1)$ and $B(G_2) > 1$. In this case, glue a block-leaf at maximum distance from S_2 to any block-leaf in G_1 to create H .

Suppose $H_2 \cap G_1 \neq \{v\}$. $H_2 \cap G_1$ is a connected graph: if not, then because H_2 is connected there is a path between any two disconnected components of $H_2 \cap G_1$ within G_2 . But any such path must begin and end at v , and therefore $H_2 \cap G_1$ itself was connected.

Thus there are at least two block-leaf vertices in $H_2 \cap G_1$, hence at least one block-leaf in $H_2 \cap G_1$ not equal to v . Choose one such vertex and label it w . Let R be the set of vertices in H_2 such that $\phi(R) = S_2$. Since $b_H(v) = 2 < B(G_2)$, by Observation 1, we must have $R = S_2$. Then $d(w, R) = d(w, S_2) > d(v, S_2)$ a contradiction. Thus $H_2 \cap G_1 = \{v\}$, and the decomposition is unique.

Case B: $B(G_2) = B(G_1)$. In this case, glue a block-leaf in G_2 to any vertex in S_1 . Since $b_H(v) > B(G_2)$, Observation 1 implies that $H_1 \cap H_2 = \{v\}$. Therefore each component of $H \setminus \{v\}$ must be entirely contained within H_1 or H_2 . Now we use that $|V(G_2)| \geq |V(G_1)|$ to conclude H is uniquely decomposable.

Case C: $B(G_2) = 1$, $B(G_1) = 0$ and $G_1 \neq K_2$. Let the *edge-distance* d' between an edge $e = \{u, v\}$ and a set of vertices S be the sum of distances between S and u and v :

$$d'(e, S) = d(u, S) + d(v, S).$$

In this case, glue an edge of G_1 to an edge $\{u, v\}$ of G_2 of maximum distance from S_2 . Suppose $G_1 \cap H_2 \neq \{u, v\}$ and let e' be an edge of $G_1 \cap H_2$ which is not e . Let R be the set of vertices in H_2 such that $\phi(R) = S_2$. Since G_1 is 2-connected, $b_H(u) = b_{G_2}(u)$ (and similarly for v), and thus $R = S_2$. Then $d'(e', R) = d(e', S_2) > d(e, S_2)$, a contradiction.

Case D: $B(G_2) = 1$, $D(G_2) > 2$ and $G_1 = K_2$. In this case, if G_2 contains a vertex of degree one, glue a vertex of G_1 to a leaf at maximum distance from S'_2 , the set of vertices of G_2 of maximum degree. Let w be the vertex in G_1 not glued to G_2 . Note that

$d(w, S'_2) = d(v, S'_2) + 1$ which is strictly greater than the distance from x to S'_2 for any leaf $x \in G_2$. Thus $w \notin H_2$ and we conclude the decomposition of H is unique.

If G_2 does not contain any vertices of degree one, then glue any vertex of G_1 to any vertex of G_2 . Then w , the vertex in G_1 not glued to G_2 , must be in H_1 and we conclude the decomposition of H is unique.

Case E: $G_2 = P_k$, $k > 3$, and $G_1 = K_2$. In this case, glue a vertex of K_2 to the third vertex along the path P_k . It is clear that this graph is uniquely decomposable. \square

In fact, this construction covers almost all uniformly-distributed two-component graphs. We fully characterize the distributions of two-component graphs by combining Theorem 19 with some examination of a few special cases.

Theorem 23. *For every graph A with connected components $G_1 \neq G_2$*

- *If neither G_1 nor G_2 is a single vertex, and $\{G_1, G_2\} \neq \{P_1, P_2\}, \{P_1, P_3\}$ (where P_i is the path with i edges), $N(A)$ is $2^{-\Omega(n)}$ -close to uniformly distributed in $G(n, p)$ modulo any q for sufficiently large p .*
- *If $A = P_1 \sqcup P_2$, $N(A)$ is $2^{-\Omega(n)}$ -close to uniformly distributed modulo q for sufficiently large p*
- *If $A = P_1 \sqcup P_3$, $N(A)$ is $2^{-\Omega(n)}$ -close to uniformly distributed modulo q if and only if q is odd. If q is even, $N(A)$ is $2^{-\Omega(n)}$ -close to being*

$$P(N(A) \equiv 2i) = 3/2q$$

and

$$P(N(A) \equiv 2i + 1) = 1/2q$$

for all $i \in \{0, \dots, q/2\}$.

- *If, without loss of generality, $G_1 = K_1$, $N(A)$ is $2^{-\Omega(n)}$ -close to being*

$$P(N(A) \equiv il) = l/q,$$

where $l = (q, n - |V(G_2)|)$

Proof. The first item follows directly from Theorem 19, Theorem 14, and Corollary 11. It remains to analyze $P_1 \sqcup P_2$, $P_1 \sqcup P_3$, and $K_1 \sqcup G_2$. We do these by direct computation.



First consider the case $A_1 = P_1 \sqcup P_2$. All gluings $H \in \mathbf{H}$ of P_1 and P_2 are illustrated above. Note that $H_2 = P_2$, $s(H_1) = 2$, $s(H_2) = 2$, $s(H_3) = 3$, and $s(H_4) = 3$. Therefore we have

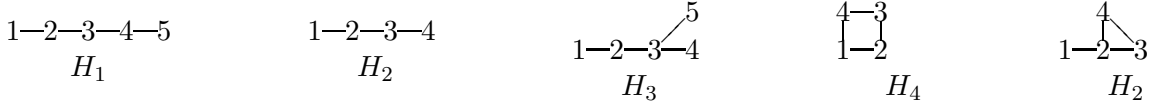
$$N(A_1) = (N(P_1) - 2)N(P_2) - 2N(H_1) - 3N(H_3) - 3N(H_4).$$

By Theorem 10, we know that the tuple $(N(P_1), N(P_2), N(H_1), N(H_3), N(H_4))$ is $2^{-\Omega(n)}$ -close to being uniformly distributed over \mathbf{Z}_q^5 . Therefore $N(A_1)$ itself is $2^{-\Omega(n)}$ -close to being uniformly distributed.

Now consider the case $A_2 = P_1 \sqcup P_3$. All gluings $H \in \mathbf{H}$ of P_1 and P_3 are illustrated below. Note that $H_2 = P_3$, $s(H_1) = 2$, $s(H_2) = 3$, $s(H_3) = 2$, $s(H_4) = 4$, and $s(H_5) = 2$. Therefore we have

$$N(A_2) = (N(P_1) - 3)(N(P_3)) - 2N(H_1) - 2N(H_3) - 4N(H_4) - 2N(H_5).$$

Again, Theorem 10 and some basic modular arithmetic are enough to generate the distributions modulo q in each case.



Finally, consider the case $A_3 = K_1 \sqcup G_2$. It is clear that $N(A) = (n - |V(G_2)|)N(G_2)$. Once more, Theorem 10 and some basic modular arithmetic are enough to generate the distribution modulo q .

□

4.2 Tree-like gluings

Graphs with more than two components are harder to work with using the methods of the previous section. As the number of components increases, the possible gluings and decompositions also increase. Nevertheless, there are some families of multi-component graphs that admit a recursive construction.

Theorem 24. *If $A = \sqcup_{i=1}^k G_i$ and there do not exist $i \neq j$ such that $G_i \subset G_j$, then there exists H_A a tree-like gluing of $\{G_i\}$ such that $H_A \neq G_i$, and $N(A)$ is $2^{-\Omega(n)}$ -close to being uniformly distributed modulo q for all q .*

Proof. Let $u_i \neq v_i$ be arbitrary vertices of G_i . Then let H_A be the graph constructed by gluing v_i to u_{i+1} . This graph is clearly tree-like, in fact it is path-like.

It is also uniquely decomposable, by induction: Suppose this construction is uniquely decomposable for all $k < n$. Now consider H_A for $k = n$. Any valid decomposition of H_A must preserve G_n (otherwise, if some vertex of G_n corresponds to a vertex of H_i , then $G_n \subset G_i$ or $G_i \subset G_n$). Therefore H_A/G_n is the construction for $n - 1$, which by hypothesis is uniquely decomposable.

Because H_A is tree-like and uniquely decomposable, by Theorem 14 $N(A)$ is $2^{-\Omega(n)}$ -close to uniformly distributed modulo q . \square

Corollary 25. *If $A = \sqcup_{i=1}^k G_i$ and the G_i are two-connected, then $N(A)$ is $2^{-\Omega(n)}$ -close to being uniformly distributed modulo q for all q .*

Proof. Any pair of two-connected graphs are incomparable. Follows from Theorem 24. \square

The reader can generate more corollaries of Theorem 24, using other families of incomparable graphs. For example, conditions on degree can guarantee incomparability.

4.3 No generic gluing exists

The previous constructions used a recursive process to create a uniquely decomposable H for G satisfying certain conditions. A natural goal would be to find a recursive process to generate H for arbitrary G . However, no such construction exists.

Theorem 26. *There does not exist a recursive construction algorithm C that, for all k and distinct G_1, \dots, G_k , generates a uniquely decomposable H_k . That is, there does not exist an algorithm C that constructs uniquely decomposable H_k by first calling C on G_1, \dots, G_{k-1} to generate H_{k-1} , and then calling C on H_{k-1}, G_k .*

Proof. Suppose there did exist such a recursive C . Let $C(G_1, \dots, G_{k-1}) = H_{k-1}$. If G_1, \dots, G_{k-1} can be glued together as a proper subgraph of H_{k-1} , then C cannot construct a uniquely decomposable H_k on input $G_1, \dots, G_{k-1}, H_{k-1}$. We note that, for example, $G_1 \subseteq G_2$ is enough to give that G_1, \dots, G_{k-1} can be glued together as a proper subgraph of H_{k-1} . \square

5 Open questions

There are two main open questions. What graphs are distributed uniformly? What families of graphs $\{G_i\}$ are uniquely or tree-like composable?

Theorem 11 gives us one means of studying graph distributions. However, it is not the case that graphs are uniformly distributed exactly when they have tree-like compositions. (Recall that $P_1 \sqcup P_2$ is uniform but is not uniquely composable.) It is possible that a more sophisticated analysis of the formula in Theorem 8 could give a different sufficient condition for graphs to be uniformly distributed.

We have fully characterized the two-component graphs that are uniquely composable, and hence admit tree-like compositions. We believe an approach similar to the two-component construction given here also works for the three-component case. However, increasing the number of components significantly complicates the analysis, and the number of cases is over twenty. We are currently developing a simpler construction for three components.

Of course, the ultimate goal is to completely characterize the uniquely composable and tree-like composable graphs with any number of components. We suspect that many, if not all, graphs admit such compositions. Theorem 26 indicates a recursive approach does not work in general, but a different type of algorithm may succeed. Even a non-constructive proof of the existence of uniquely decomposable or tree-like graphs would be interesting.

References

- [1] Frank Harary: *A survey of the reconstruction conjecture*, Graphs and Combinatorics (Ruth Bari and Frank Harary, eds.), Lecture Notes in Mathematics, vol. 406, Springer Berlin / Heidelberg, 1974, pp. 18-28.
- [2] P. G. Kolaitis and S. Kopparty, *Random graphs and the parity quantifier*, arXiv.org:0904.2436, 2009.